# *If*-conditionals in ICLE and the BNC:
# A success story for teaching or learning?

Costas Gabrielatos
Edge Hill University

## Abstract

This paper aims to contribute to the methodological toolbox of "pedagogy-driven corpus-based research" (Gabrielatos 2006), that is, research which is situated at the intersection of language description, pedagogical lexicogrammar, and pedagogical materials evaluation (*e.g.* Harwood 2005; Hunston & Francis 1998; Kennedy 1992; Owen 1993). The contribution of the present paper mainly lies in proposing a method of triangulating the corpus-based evaluation of lexicogrammatical information in English as a foreign language (EFL) coursebooks, by way of examining a relevant corpus sample of learner written output.[1]

## 1. Motivation, background and aims

An increasing body of corpus-based research comparing the information in English language teaching (ELT) coursebooks and other pedagogical materials has provided compelling indications that pedagogical information is too often at odds with the evidence in L1 corpora − see Meunier & Gouverneur (2009: 181-186) for a survey. Regarding conditionals, Gabrielatos (2003) examined the relevant information and examples in fifteen coursebooks for learners of English as a foreign language (EFL) − eleven at intermediate level and four at advanced level. The pedagogical information was contrasted to a random sample of 781 *if*-conditionals from the written BNC (Aston & Burnard 1998), derived using *BNCweb* (see Hoffmann *et al.* 2008). The comparison was repeated using the same sample from the written BNC (henceforth BNCw), but examining eleven coursebooks at advanced level (Gabrielatos 2006).

---

[1] This is a revised and expanded version of my presentation at LCR2011 (Gabrielatos 2011), which also reports on additional analysis carried out after the conference.

Both analyses (Gabrielatos 2003, 2006) revealed that the pedagogical information in the coursebooks, taken collectively, presented learners not only with a partial picture of the variety of types of conditionals and their respective morphosemantic features, but also a distorted one.[2] Table 1 below indicates the proportion of conditionals in the BNCw sample that conform to the following:

a. The *traditional* typology (essentially the one used in Logic − see also Leech 2011: 17): *First* (*If* + Present Simple → *will* + inf.), *Second* (*If* + Past Simple → *would* + inf.) and *Third* (*If* + Present Perfect → *would* + Perfect inf.).

b. The typology that would emerge by taking into account the information shared by the coursebooks in the samples − termed *consensual.*

c. The typology that would emerge by collating the information given in all coursebooks in the samples − termed *inclusive*.

|  | Traditional | Consensual | Inclusive |
|---|---|---|---|
| Intermediate + Advanced | *14.9%* | *29.3%* | *43.6%* |
| Advanced | *13.8%* | *27.8%* | *72.5%* |

*Table 1. Coverage of pedagogical information on conditionals*

Clearly, even the advanced coursebooks, which would be expected to contain more comprehensive and nuanced information, seriously under-represent the variety of types and morphosemantic attributes of *if*-conditionals found in BNCw. Even the 'inclusive' typology in the case of advanced coursebooks would fail to account for more than one in four (27.5%) of the *if*-conditionals in the BNCw sample (Table 1 above). In fact, this higher representation is due to the information in a single coursebook. The observed under-representation mainly hinged on the issues outlined below.

The coursebook typologies ignore the type of conditionals termed "indirect" (Quirk *et al.* 1985), "speech act" (Sweetser 1990), or "pragmatic" (Athanasiadou & Dirven 1997), and

---

[2] For an outline of different permutations of the ELT typology, see Gabrielatos (2006).

only explicitly focus on "direct" conditionals (Quirk *et al.* 1985).[3] In both types, as in all conditionals, the *protasis* "forms a background against which comments are proffered" in the *apodosis* (Sweetser 1990: 125). The *protasis* is the part in which a condition is specified, whereas the *apodosis* is the part which specifies the action, event, (change of) state, conclusion or speech act that is somehow contingent on the condition. The nature of this contingency determines whether a conditional is direct or indirect.

In direct conditionals, the realisation of the content of the apodosis depends on the realisation, actuality or factuality of the content of the protasis. For example, in (1) below, the prediction expressed in the apodosis is directly dependent on the actualisation of the premise in the protasis. That is, changes in the premise (*i.e.* the content of the protasis) are expected to result in changes in the conclusion (*i.e.* the content of the apodosis). If the package of cash can be assembled, then there is a likelihood that the key manager will be convinced; if the package cannot be assembled, then the manager cannot be expected to be convinced.

(1) I know they're asking more than we want to pay, but if we can assemble a package of cash, stock options, and newly issued shares as a good inducement, I think we'll convince the key manager and he'll persuade the others to sell. [FPB 108]

In indirect conditionals, what is contingent on the protasis is not the content of the apodosis, but either the relevance of its very uttering, or the wording or clarity of its content – exemplified by (2) and (3) respectively.

(2) If antibiotics are likely to clear up the infection, why are we having this long discussion? [CH1 5292]

(3) He's not a bad sort, for a brother if you know what I mean [AN7 3257]

In (2), positing the question in the apodosis becomes relevant only if the content of the protasis is actual/true. In (3), the actuality of the protasis does not affect the meaning expressed in the apododis, or the speech act performed through it; rather, it invites the listener to seek clarification if the content of the apodosis is not clear to them.

The fundamental shortcomings of the ELT typology are its two central classification criteria. First, ELT types are classified according to the degree of likelihood expressed in the apodosis,

---

[3] Please note that the first three terms do not refer to exactly the same group of conditionals, although the groups they define overlap to a great extent. For a critical examination of different classifications of English conditionals in linguistics, and a corpus-based adaptation of the typology in Quirk *et al.* (1985), see Gabrielatos (2010, Chapters 4 and 6).

with apodoses expressing other modal senses (*e.g.* obligation − as in (4) below) being ignored (see Gabrielatos 2006, 2010: 230-265).

> (4) If he wants to get or to keep his employment, he has to sign the document which the employer puts before him. [J7B 92]

Although coursebooks may provide examples of conditionals with apodoses expressing other modal notions, such examples are a minority, and, more importantly, they are not linked to types explicitly specified in the ELT typology. Second, ELT types are distinguished according to their time reference − a classification criterion which is absent in linguistic typologies, and which has been explicitly deemed unhelpful (Huddleston & Pullum 2002: 738).

The above shortcomings are most clearly evident in the problematic distinction between *Zero* and *First* conditionals.[4] The *Zero* type is presented as expressing general truths, or cause-effect relations that always pertain − in contrast to the *First* type, which is presented as referring to specific situations that are likely to pertain in the future. However, lack of modal marking in the apodosis does not ensure that the conditional expresses a general/timeless relation, nor does presence of modal marking ensure that the conditional refers to a specific future situation. Therefore, conditionals conforming to the form of the *Zero* type can also express specific conditional relations with a future time reference, whereas conditionals conforming to the form of the *First* type can also express general conditional relations with a timeless reference − as Table 2 demonstrates (adapted from Gabrielatos 2006).

| | General / Timeless | Specific / Future |
|---|---|---|
| *Zero* | The argument obviously generalises to show that, if there **is** a non-negative solution of (9.8) with <gap desc=formula>, then any new tableau obtained by pivoting in column j **is** efficient. [CA4 738] | If Bridges **is** right, this still **does not avoid** possible legal argument over the "reasonableness" of the contract between purchaser and provider, nor over how well contracts are complied with. [CR5 693] |
| *First* | If a Troll **suffers** harm his flesh **will** almost instantly **re-grow**. [CMC 250] | "If they **charge** the wrong man, it**'ll make** a difference to him!" said Melissa dryly. [HNJ 1807] |

*Table 2. Zero and First conditionals: two sides of the same coin*

---

[4] Typically, their form is presented as '*If* + Present Simple → Present Simple' and '*If* + Present Simple → *will* + infinitive' respectively.

In addition, ELT coursebooks generally adopt a naïve and restricted approach to modal marking, as they tend to focus on central modals[5] (Gabrielatos 2003, 2006). This is misleading, as well as potentially confusing and restricting for learners, as it fails to account for almost one-third (32.7%) of the modal markers in the apodosis of the conditionals in the BNCw sample, and almost half (48.2%) of all modal markings. Table 3 shows the types of other modal markers found in the sample (either in the protasis or the apodosis).[6]

| Category | Examples |
| --- | --- |
| Marginal modals | *dare, need, ought to, used to* |
| Modal idioms | *had better*, *have got to* |
| Semi-auxiliaries | *have to, be able to, be going to, be obliged to* |
| Catenative verbs | *appear to, claim to, intend to, seem to, tend to* |
| Other constructions involving verbs (including mental state predicates) | *think that, know that, believe that, doubt that, be sure that* |
| Constructions involving nouns | *the possibility exists that, be under an obligation to* |
| Constructions involving adjectives | *it is possible that, it is desirable to* |
| Constructions involving adverbs | *potentially, probably* |
| Imperative | *Come in!* |
| Past tense | *If they changed their minds …* |
| Present/Past Subjunctive | *If it were possible …* |

*Table 3. Modal markers other than central modals*

The morphosemantic specifications of the ELT types (and the examples given) under-represent the extent of modal marking in protases − particularly in *First* conditionals. Overall, the information and examples communicate (explicitly or implicitly) that, a) protases with modal marking are rare, and b) if the protasis or apodosis is modally marked, this is done through a single modal marker − that is, the possibility of multiple modal markers in one or both of the parts of a conditional is not mentioned (Gabrielatos 2003, 2006). However, this is not supported by empirical evidence. In the BNCw sample used in Gabrielatos (2010), 41.7% of protases are modally marked, and 20.8% of apodoses have two or more modal markers − as demonstrated by (5):

---

[5] Central modals are: *can, could, may, might, must, shall, should, will, would*.
[6] For details, see Quirk *et al.* (1985: 136-148) and Gabrielatos (2010: 119-125, 148-151).

(5) If you should decide to concentrate on one particular nursing specialty then you **will probably want** to undertake a clinical nursing studies course. [CHT 248]

Finally, the morphosemantic attributes presented as 'exceptions' or 'special cases' in the coursebooks proved to be too frequent to be accurately described as such. A case in point is Past tense marking with past time reference in protases. The coursebooks in the sample either ignore this use, or present it as a special case, stressing the epistemic interpretation of Past tense marking as the default one. However, in the BNCw sample, more than one-third (34.6%) of Past tense marking in protases expressed past time − as in (6) below.

(6) After a decade of being made to feel that you were somehow lacking as a person **if** you **didn't manage** to cram thrice-weekly workouts into your schedule, it was high time that a good reason not to exercise emerged. [AK6 1119]

On the basis of the above results, it was hypothesised that learner written production –when compared to similar texts in the written BNC – would be characterised by the following:

Regarding types of conditionals:

    a.  Under-representation of indirect conditionals.

    b.  Among direct conditionals, an over-representation of conditionals with apodoses expressing degrees of likelihood, and a corresponding under-representation of other types.

    c.  A higher proportion of *Zero* types in ICLE, that is, *if*-conditionals without additional[7] modal marking in their protases and apodoses.

Regarding modal load and particular modal markers:

    d.  Smaller modal load in *if*-conditional constructions, as well as in protases and apodoses.

    e.  Lower proportion of Past tense marking with past time reference in protases.

    f.  Under-representation of central modals in protases.

    g.  Over-representation of central modals in apodoses.

The remainder of the paper will be concerned with testing the above hypotheses using a corpus-based approach.

---

[7] The word 'additional' is used because the word *if* has already marked the construction for modality (see Gabrielatos 2010: 148-149, 185, 296-301).

## 2.    Data, metrics and methodology

The main methodological contribution of this study is in proposing an approach to triangulating results derived from different types of data within the framework of pedagogy-oriented frequency studies (Figure 1) − see Leech (2011) for an overview.
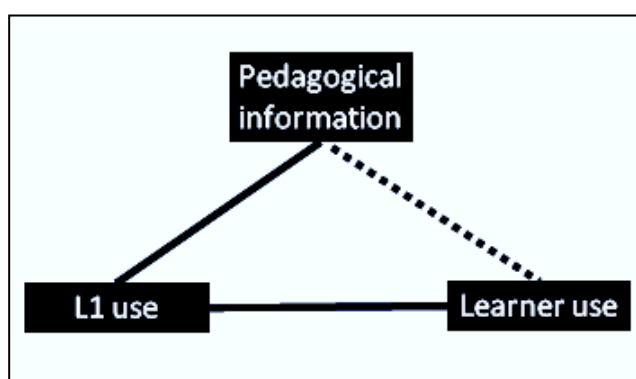


*Figure 1. Triangulating pedagogy-oriented frequency studies*

The comparison between L1 use and learner use offers insights into the language features (or the attributes and uses of a particular language feature) that learners may have problems in using, and, therefore, suggests areas that language instruction can usefully focus on. The comparison between pedagogical information and L1 corpus evidence provides insights into the level of representativeness of the former, and, therefore, suggests areas in which pedagogical information needs to improve. Finally, the comparison of pedagogical information with learner production can suggest the extent to which the former affects the latter. However, the inference of such a causal connection must be treated with caution,[8] for two main reasons (see also Section 4):

    a.  the content of the pedagogical materials examined in a given study may diverge from that of the materials used with the learners whose output is represented in the corpus used in the study.

    b.  the learners' sources of input need not be restricted to pedagogical materials.

This study uses two random samples of *if*-conditionals representing learner and L1 use in written discourse. The former is extracted from a corpus of learner writing at upper-intermediate and advanced levels, namely ICLE (Granger *et al.* 2002), the latter through

---

[8] The dotted line in Figure 1 above represents the tentative nature of any causal links between pedagogical information and learner use.

BNCw. However, ICLE only contains argumentative essays, whereas BNCw is richer in text types. For reasons of comparability, only the instances from academic texts, essays and editorials in the BNCw sample were considered (BNCaee) – resulting in a sample of 195 *if-*conditionals. Therefore, a random sample of 200 instances was drawn from ICLE, using *CQPweb* (Hardie 2012) and a complex query developed in Gabrielatos (2010)[9] – resulting in 190 true instances of *if*-conditionals. Each *if*-conditional was manually annotated for its type, according to a) the typology used in ELT, and b) the typology developed in Gabrielatos (2010), as well as the modal marker and the type of modality marked in the protasis and apodosis of each *if*-conditional in the samples. Frequency differences are also tested for their statistical significance.[10] Due to the small size of the samples, the threshold of $p{\leq}0.05$ was used.

The extent of modal marking in the two samples will be measured using a metric developed in Gabrielatos (2010: 50-52), namely, *modal load*. The modal load is best seen as a two-dimensional metric, as it is established through the interaction of two complementary metrics: *modal density* and *modalisation spread* (see Section 3, Figure 2). Modal density is the average number of modal markings per clause, and is expressed as the number of modal markings per hundred clauses. Modalisation spread is the proportion of constructions that carry at least one modal marking, and is expressed as the proportion (%) of modalised constructions in a sample. Modal density helps comparisons between samples by normalising for the complexity of the constructions in each (see Ball 1994: 297-300); modalisation spread corrects for heavily modalised constructions in the sample.[11]

## 3.    Analysis and discussion

A first interesting observation, not anticipated in the hypotheses, is that ICLE contains about a quarter (23.5%) more *if*-conditionals than BNCaee, and the difference has a very high statistical significance ($p{\leq}10^{-17}$).[12] This may be due to the prominence of conditionals in ELT materials, and the attendant time spent on input and practice, in that learners may feel that they are expected to regularly include conditionals in their output. Of course, as ELT materials predominantly focus on *if*-conditionals, it is also plausible that learners may tend to

---

[9] The query has a precision of about 95%.

[10] The calculations were carried out with Paul Rayson's online log-likelihood calculator: http://ucrel.lancs.ac.uk/llwizard.html.

[11] For adaptations and other applications of the two-dimensional metric, see Gabrielatos *et al.* (2010) and Torgersen *et al.* (2011).

[12] Relative frequencies are 2480.2 and 2008.7 per million words, respectively –extrapolated from the query results and precision.

underuse other conditional constructions. In the remainder of this section we will examine the hypotheses outlined in Section 1.

Despite the absence of indirect conditionals in the ELT typology, the ICLE sample contains an almost identical proportion to that in BNCaee (3.2% and 3.6% respectively). This may well be due to ICLE learners having received input outside teaching materials. However, it can also be argued that academic writing is not a suitable environment for the examination of indirect conditionals, as they are much more frequent in spoken English (Gabrielatos 2010: 268-270).

What seems to have a clear influence on students' written production of *if*-conditionals is the structuring of the ELT typology around the degree of likelihood expressed through them. The ICLE sample contains 13.5% more *if*-conditionals with apodoses expressing likelihood, and 40.6% fewer *if*-conditionals expressing deontic notions when compared to the BNCaee sample (in both cases, $p \leq 0.05$) − exemplified by (7) and (8) respectively.

(7) If these poor people work and work and work and are not respected like such rich people they may slip into the criminal scene. [GEDR 1013]

(8) If a state eagerly wants to kill people in a legal way, the government should create a law to do that in an instant and not after 25 years. [GESA 4011]

When looking at the proportion of *Zero* conditionals, the frequency difference is not particularly large, nor statistically significant: the ICLE sample contains about 20% more *Zero* conditionals than the BNCaee sample. Given the tenuous distinction between *Zero* and *First* demonstrated above, it would be useful to conflate the four permutations exemplified in Table 2, and compare their collective frequency in the ICLE and BNCaee samples. At the same time, an approach will be adopted which takes into account all instances of modal markers (not just central modals) and which makes no typological distinctions in terms of time reference. In fact, the resulting composite type is recognised in two linguistic typologies of conditionals, and has been termed an *Open* conditional (Huddleston & Pullum 2002: 738, Quirk *et al*. 1985: 1091). The comparison reveals that the ICLE and BNCaee samples contain comparable proportions of *Open* conditionals: 75.8% and 77.9% respectively. This suggests one or more of the following: a) ICLE learners use *if*-conditionals more flexibly than a strict adherence to the ELT typology would predict, b) the instructional input they have received was unusually comprehensive, c) they had contact with English outside of formal instruction.

Tentative indications of the effect of the information in the coursebooks on learner output are given by the examination of the modal load of the *if*-conditional constructions in the two samples, as well as the modal load of protases and apodoses taken separately. The ICLE sample shows lower modal density and modalisation spread than the BNCaee sample in all comparisons (Figure 2).
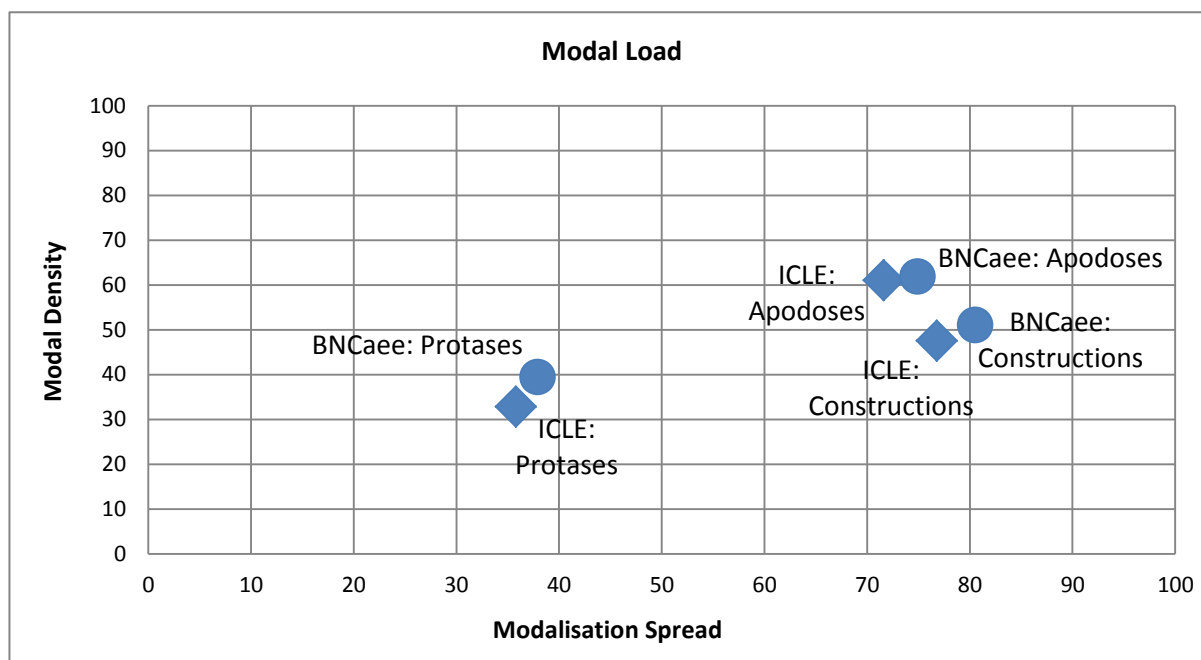


*Figure 2. ICLE and BNCaee: Modal Load comparison*

However, these differences are neither particularly large, nor statistically significant. In the case of whole constructions and apodoses, the differences are between 1.5% and 7.0%. Larger differences, but still of relatively small size, are seen in the modal load of protases, with ICLE showing 12.8% lower modal density and 12.7% lower modalisation spread. The emerging picture is, therefore, that only the modal load of protases seems to be affected by the information in the coursebooks – but even in this case, the influence is fairly mild. Much more pronounced, but still not statistically significant, is the frequency difference in the use of the Past tense in protases with modal rather than temporal sense, with the ICLE sample having 42% more instances of the modal sense than the BNCaee sample.

Finally, no difference emerges from the comparison of the proportion of central modals in protases and apodoses, with the ICLE and BNCaee samples showing almost identical proportions (Table 4).

|  | ICLE | BNCaee |
|---|---|---|
| Protases | 17.3% | 21.0% |
| Apodoses | 69.9% | 69.5% |

*Table 4. Proportions of central modals*

However, the similar frequencies may hide differences in the variety of modal markers other than central modals in the two samples. This can be examined through the type-token ratio. The type-token ratio values for non-central modal markers, for ICLE and BNCaee respectively, are 0.33 and 0.44 in protases, and 0.57 and 0.67 in apodoses. That is, the ICLE sample is fairly distinctly less rich in its variety of modal markers.

## 4.     Conclusions and further research

Overall, the comparisons have shown trends compatible with all hypotheses presented in Section 1. However, if we take into account the sizes of the differences, and the fact that only one difference was statistically significant, it would seem that the overall use of *if*-conditionals by the ICLE students is not strongly influenced by the shortcomings of the ELT typology and related pedagogical information (outlined in Section 1). Only hypothesis (b) is strongly supported by the results, being the only difference that was both sizeable and statistically significant. This difference, however, is particularly important, as it pertains to a criterion at the heart of the ELT typology: the use of the degree of likelihood as the sole major feature distinguishing between types. A second large (but not statistically significant) difference is the underuse in ICLE of the Past tense with past time reference in protases. Finally, ICLE learners seem to use a smaller variety of modal markers compared to BNCaee. In all other cases, the comparisons yielded fairly similar results. Despite the latter observation, it would be premature to conclude that the explicit information in coursebooks has limited influence on learner production, for the reasons outlined below.

In light of most results being statistically non-significant, one explanation for the findings would be that differences do exist, but the corpus samples were too small for the observed differences to be dependable. In terms of corpus selection, we need to consider that, apart from the distinction between native and non-native production, ICLE and the BNC differ in the level of expertise of the writers whose production is included in the corpora. The BNC

contains texts by individuals who write as part of their profession (*e.g.* reporters, novelists, academics), whereas ICLE contains the output of novice writers (see Gabrielatos & McEnery 2005: 312-313). In addition, the BNC contains texts that have been edited and proofread by persons other than the writer, whereas this is not the case with ICLE texts. Therefore, the written BNC (irrespective of sub-corpora) may not be the most appropriate reference corpus.

As regards language teaching and learning, it would be useful to keep in mind that ICLE contains the written production of learners from a variety of educational contexts. It is not unlikely that pedagogical materials may have been employed, the content of which diverges from those examined in Gabrielatos (2003, 2006). It is equally likely that, whatever the teaching materials, teachers adapted and/or supplemented the pedagogical information in them. A further useful consideration is that what is explicitly taught is not necessarily, or solely, what is learned. That is, the results may be attributable more to the learners' own learning strategies, or their contact with English outside the classroom, than to the information in ELT materials, or explicit teaching in general. ICLE learners may have derived information on *if*-conditionals implicitly through the texts in the coursebooks, which may well contain conditionals not covered by the ELT typology. They may also have had occasion to learn on their own – an eventuality which is particularly likely given the increasing online availability of spoken and written English texts, as well as online opportunities to interact in English. To put it bluntly, we may be wise to ask whether the use of *if*-conditionals in ICLE and BNCaee mostly shows similarities or fairly small differences *because* of the content of ELT materials or *in spite* of it.

Finally, ICLE learners have a variety of L1s and, therefore, the possibility cannot be discounted that the picture emerging from the present analysis may hide L1-specific variation. A complementary line of enquiry would be to examine the extent to which learners use ELT types consciously, that is, whether their production or self-editing is guided by explicit pedagogical information. Similarly, it would be useful to establish whether learners use modal markers other than those presented in ELT materials (*e.g. want, be obliged to*) in awareness of their modal sense. Of course, such an investigation would need to encompass research techniques outside of corpus linguistics (see Hollmann & Siewierska 2006).

Therefore, apart from examining larger samples, the following can be expected to provide clearer and more reliable indications as to the extent of the influence that pedagogical

information may have on learner output. The *if*-conditionals in ICLE would be better compared with those in a corpus of novice L1 writing elicited under conditions similar to ICLE (*e.g.* LOCNESS).[13] In the same vein, the comparison of samples from spoken L1 and learner corpora will provide better indications regarding the influence of pedagogical materials on the proportion of indirect conditionals in learner production. Finally, comparisons of *if*-conditionals produced by learners with different L1s and/or learners from different countries can shed light on the influence of L1 interference or the effect of educational context.

## References

Aston, G. & Burnard, L. (1998). *The BNC Handbook: Exploring the British National Corpus*. Edinburgh: Edinburgh University Press.

Athanasiadou, A. & Dirven, R. (1997). Conditionality, hypotheticality, counterfactuality. In A. Athanasiadou & R. Dirven (eds) *On Conditionals Again*. Amsterdam: John Benjamins, 61-96.

Ball, C.N. (1994). Automated text analysis: Cautionary tales. *Literary and Linguistic Computing* 9(4), 265-302.

Gabrielatos, C. (2003). Conditional sentences: ELT typology and corpus evidence. Paper given at the *36th Annual BAAL Meeting*, University of Leeds, UK, 4-6 September 2003. [http://eprints.lancs.ac.uk/140]

Gabrielatos, C. (2006). Corpus-based evaluation of pedagogical materials: *If*-conditionals in ELT coursebooks and the BNC. 7th *Teaching and Language Corpora* Conference, University Paris 7 Denis Diderot, 1-4 July 2006. [http://eprints.lancs.ac.uk/882]

Gabrielatos, C. (2010). *A corpus-based examination of English* if-*conditionals through the lens of modality: Nature and types*. Doctoral dissertation, Lancaster University, UK. [http://ethos.bl.uk/OrderDetails.do?did=1&uin=uk.bl.ethos.539699]

Gabrielatos, C. (2011). *If*-conditionals in ICLE and the BNC: A language teaching success story? *Learner Corpus Research 2011*, Louvain-la-Neuve, Belgium, 15-17 September 2011. [http://eprints.lancs.ac.uk/50096]

Gabrielatos, C. & McEnery, T. (2005). Epistemic modality in MA dissertations. In P.A. Fuertes Olivera (ed.) *Lengua y Sociedad: Investigaciones recientes en lingüística*

---

[13] http://www.uclouvain.be/en-cecl-locness.html.

*aplicada*. Lingüística y Filología 61. Valladolid: Universidad de Valladolid, 311-331. [http://repository.edgehill.ac.uk/4137]

Gabrielatos, C., Torgersen, E., Hoffmann, S. & Fox, S. (2010). A corpus-based sociolinguistic study of indefinite article forms in London English. *Journal of English Linguistics* 38(4), 297-334.

Granger, S., Dagneaux, E. & Meunier, F. (2002). *The International Corpus of Learner English. Version 1.1*. Handbook and CD-ROM. Louvain-la-Neuve: Presses Universitaires de Louvain.

Hardie, A. (2012). CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 17(3), 380-409.

Harwood, N. (2005). What do we want EAP teaching materials for? *Journal of English for Academic Purposes* 4(2), 149-161.

Hoffmann, S., Evert, S., Smith, N., Lee, D. & Berglund-Prytz, Y. (2008). *Corpus Linguistics with BNCweb - A Practical Guide.* Frankfurt am Main: Peter Lang.

Hollmann, W. & Siewierska, A. (2006). Corpora and (the need for) other methods in a study of Lancashire dialect. *Zeitschrift für Anglistik und Amerikanistik* 54, 203-216.

Huddleston, R. & Pullum, G.K. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Hunston, S. & Francis, G. (1998). Verbs observed: A corpus-driven pedagogic grammar. *Applied Linguistics* 19(1), 45-72.

Kennedy, G. (1992). Preferred ways of putting things with implications for language teaching. In J. Svartvik (ed.) *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82, Stockholm, 4-8 August 1991*. Berlin: Mouton de Gruyter, 335-378.

Leech, G. (2011). Frequency, corpora and language learning. In F. Meunier, S. De Cock, G. Gilquin & M. Paquot (eds) *A Taste for Corpora: In Honour of Sylviane Granger*. Amsterdam & Philadelphia: John Benjamins, 7-31.

Meunier, F. & Gouverneur C. (2009). New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material. In K. Aijmer (ed.) *Corpora and Language Teaching*. Amsterdam & Philadelphia: John Benjamins, 179-201.

Owen, C. (1993). Corpus-based grammar and the Heineken effect: Lexico-grammatical description for language learners. *Applied Linguistics* 14(2), 167-187.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Sweetser, E.E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press.

Torgersen, E., Gabrielatos, C., Hoffmann, S. & Fox, S. (2011). A corpus-based study of pragmatic markers in London English. *Corpus Linguistics and Linguistic Theory* 7(1), 93-118.